

AD-A039 707

NAVAL UNDERWATER SYSTEMS CENTER NEWPORT R I
SUBJECT INDEXING -- DDC AND THE DOD TECHNICAL LIBRARIES. (U)
FEB 77 R G HEROUX

F/G 5/2

UNCLASSIFIED

NL

1 OF 1
AD
A039707



END

DATE
FILMED
6-77



2
B.S.

6

SUBJECT INDEXING -- DDC AND THE DOD TECHNICAL LIBRARIES.

ADA 039707

10

by
Ronald G. Heroux

Head of User Services
Technical Library
Naval Underwater Systems Center
Newport, R.I.

9

Doctoral thesis,

11

February 1977

12

15p.

AD No. —
DDC FILE COPY

DDC
RECEIVED
MAY 23 1977
A

(This paper was written in partial fulfillment of a Doctor of Arts degree
in Library Science at Simmons College, Boston, Mass.)

UNLIMITED DISTRIBUTION

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

406 068

1B

INTRODUCTION

The Defense Documentation Center's (DDC) indexing or cataloging information has been little used by Department of Defense (DOD) technical libraries in cataloging documents. Yet, Library of Congress' cataloging has been extensively employed by these same libraries when books are entered into the library system. Why is one national system accepted and not the other?

Since most DOD technical libraries have a substantial collection of technical documents whose contents are often very specialized, many DOD technical librarians feel that only they can satisfactorily index documents to meet the retrieval needs of the scientists and engineers working in their organizations. Some DOD librarians also feel that even if DDC's indexing were used by their libraries, it could not be applied to reports arriving on automatic distribution because of the delay in receiving DDC indexing once the report has been received in-house.

Inasmuch as the adequacy or the exhaustiveness of DDC's indexing, specifically their subject indexing, is one of the main reasons why DDC's cataloging has not been widely used by DOD technical libraries, a comparative study of DDC subject indexing with in-house indexing at DOD technical libraries across the country appears warranted.

OBJECTIVES OF THE STUDY

1. To determine whether DDC or the DOD technical libraries use more subject indexing terms per document.
2. To determine what percentage of subject indexing terms assigned by DOD technical libraries are identical or similar to those used by DDC.

White Section <input checked="" type="checkbox"/>	
Buff Section <input type="checkbox"/>	
UNCLASSIFIED <input type="checkbox"/>	
CLASSIFICATION <input type="checkbox"/>	
BY <input type="checkbox"/>	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	AVAIL. and/or SPECIAL
A	

METHODOLOGY

A statistically random sample of technical documents indexed according to subject by DDC and a selected number of technical libraries was examined. The sample was stratified according to the DOD libraries chosen. This stratification avoids an average outcome which might not be representative of the type of subject indexing performed by any of the libraries in the sample when compared to DDC. It also allows for comparisons by service branch (Navy, Army, Air Force). For each organization chosen, the population was numbered and the sample picked randomly with the use of A Million Random Digits developed by Rand Corporation in 1955.

The population from which the sample was drawn included documents conforming to the following characteristics:

1. The bibliographic and subject indexing information concerning the documents was available on the DDC RDT&E on-line system.
2. The documents were unclassified. (Confidential and secret documents were excluded.)
3. The documents were produced by one of 22 DOD organizations chosen from a random sample of 49 DOD organizations having a DDC RDT&E on-line terminal. (These 49 organizations are, in all probability, the heaviest requesters of DDC documents and probably input the vast majority of military-produced documents into the DDC system.)
4. The documents contained a 1975 or 1976 report date. (This guaranteed an examination of the current subject indexing system, for many libraries have changed their cataloging procedures with the addition of automated library operations.)

Twenty-five randomly selected documents were chosen from the organizations selected. Each library whose organization produced the reports in the sample was sent all the bibliographic information necessary to identify the documents in question, and each library was requested to forward the cataloging information (specifically the subject indexing) of these documents to this author.

Prior to requesting this information, the 22 libraries chosen were phoned, and librarians were briefed on the study. Also, five specific questions were asked concerning the practices of subject indexing in-house. (These questions are listed in Appendix A.)

The DDC subject indexing information was obtained from the DDC on-line system where the population was also derived.

To meet the first objective of this study outlined previously, the subject indexing terms assigned by DDC to the documents in the sample were counted and numerically compared with subject indexing terms used by the DOD libraries. To meet the second objective, the subject indexing terms assigned by DDC to the documents in the sample were compared with the terms used by DOD libraries to determine if the terms were identical or similar.

In order for the subject indexing terms (which could include one or more words) to be considered identical, each term had to match each other to the letter, except in cases when one library assigned the term in the singular and DDC assigned it in the plural or vice versa. These terms were also considered identical; e.g., laser=lasers, vibration test=vibrations tests, etc.

Subject indexing terms were considered similar when:

1. The DOD library used an abbreviation of the term and DDC used the term spelled out or vice versa; e.g., FFT=Fast Fourier Transform, MRE=Mean Radial Error, etc.

(4)

2. The same word(s) (either in the singular or plural) representing the concept indexed was present in both terms; e.g., calibration techniques=calibration, SSN 688 class submarines=SSN 688 vessels, etc.
3. The term (either singular or plural) assigned by the DOD library was split into more than one term by DDC or vice versa while maintaining the same words representing the concepts in the term not split; e.g., narrowband sonar=narrowband (and) sonar, Trident sonar system=Trident submarines (and) sonar, hydro-acoustic transducers=hydroacoustics (and) transducers, etc.

ANALYSIS OF FINDINGS

Of the 22 DOD libraries (see Appendix B) picked in the sample of DDC on-line users 4 were Air Force, 8 Army and 10 Navy. The sample was picked in proportion to the total number of on-line users within each branch of the service. Both in the following tables and throughout this analysis each branch of the service is treated separately when compared with DDC. This was done because there was concern that there might be some significant discrepancies among the three services which would not be noticed if they were solely compared as one group under DOD. All service branches (Army, Navy and Air Force) are also lumped together under the heading of DOD and compared with DDC as a group.

In phoning each of the 22 libraries sampled, it was discovered that several libraries were actually using the cataloging done by DDC; consequently, requesting their indexing was not necessary since they would yield a 100% matchup. Libraries in this category amounted to four: two Army, one Navy and one Air Force. In

(5)

asking the librarians interviewed in each of these four libraries why they used DDC versus in-house cataloging as the remaining 18 of their sister libraries were doing, the reason given was overwhelmingly an economic one. "Not enough staff to spend time on cataloging," one librarian said point blank.

Thus attrition combined with personnel cutbacks dictated the direct use of DDC's cataloging. There was no indication that any exhaustive analysis was undertaken before the changeover to DDC to examine whether or not a change was justified based on the quality or exhaustiveness of DDC's cataloging. Although all four librarians interviewed were pleased with the subject indexing done by DDC and felt it met the needs of their community, this study still remains important in light of the fact that only a small number of libraries (primarily the smaller ones) are using DDC's indexing information in cataloging their documents. Moreover, no recent study was discovered comparing DDC's indexing of technical reports with that of in-house indexing on the local level. (Cathy Lyon did a comparative subject indexing study of 551 Work Unit Summaries indexed at the Naval Surface Weapons Center and at DDC - "A Comparative Evaluation of Machine-aided Indexing at the Defense Documentation Center vs. the Manual Indexing done by the NSWSC/Dahlgren Laboratory Technical Library," 1975, ADA 009 237.)

Libraries Submitting Indexing

Beside the four libraries above which were not asked to submit their indexing because they already used DDC, four other libraries were also not included in the subject indexing sample. The latter four did not supply their indexing information. Thus 14 libraries submitted their indexing--four Army, nine Navy and one Air Force.

In comparing Air Force indexing with that of DDC, there is a strong bias toward the one library which submitted the indexing. Whether this Air Force

library's indexing is representative of other Air Force libraries remains uncertain. Also, there being a large number of Navy libraries biases DOD results in favor of the Navy. Nevertheless, as can be seen in the following tables, there are not many significant differences between the military services, thus the Navy bias becomes negligible.

A total of 320 documents were analyzed for subject indexing--212 from the Navy, 85 from the Army and 23 from the Air Force. A few documents were not analyzed because local indexing was not available.

Numerical Comparison

In attempting to determine whether or not DDC's subject indexing was as exhaustive as in-house indexing, two variables were measured: (a) a numerical comparison was made as well as a (b) comparison of similar and identical terms. First the former will be examined.

Table 1 clearly shows that DDC assigns two and a half times as many subject indexing terms to documents as do the DOD libraries. Except for the Air Force which DDC outnumbers 4 to 1, the DOD ratio remains about the same for the Navy and the Army.

Table 1 AVERAGE NUMBER OF SUBJECT INDEXING TERMS ASSIGNED PER DOCUMENT.

Navy.....	7	DDC.....	15
Army.....	9	DDC.....	18
Air Force.....	3	DDC.....	14
*DOD.....	6.3	DDC.....	15.3

*(Includes the average of Army, Navy and Air Force)

Except for the library at the Army Missile Command, DDC assigned more subject terms per document than any of the libraries sampled. In general, libraries having

(7)

an automated retrieval system (see Table 2), such as the Army Missile Command library, assigned more subject indexing terms per document than those libraries without an automated retrieval system.

Table 2 LIBRARIES USING AN AUTOMATED RETRIEVAL SYSTEM FOR DOCUMENTS.

(*Since all 22 libraries were initially interviewed by phone, this table represents all 22 rather than the 14 libraries which submitted subject indexing information.)

	<u>YES</u>	<u>NO</u>
Navy	7	3
Army	3	5
Air Force	1	3
DOD	11	11

Identical & Similar

Table 3 indicates that an average of only 55.7% of the terms assigned by DOD libraries are identical (or similar) to those terms assigned by DDC. The true percentage is probably more around 50% since the Air Force, represented by only one library, has a total of 66%, which may not truly be representative of the Air Force.

Table 3 ASSIGNMENT OF IDENTICAL AND SIMILAR INDEXING TERMS: COMPARISON OF DOD LIBRARIES AND DDC.

	<u>Identical to DDC</u>	<u>Similar to DDC</u>	<u>Total</u>
Navy	32%	19%	51%
Army	27%	23%	50%
Air Force	37%	29%	66%
DOD*	32%	23.7%	55.7%

(*Includes the average of Army, Navy, and Air Force.)

Although terms assigned by the DOD Libraries matched only 55.7% of the terms assigned by DDC, 31.3% (Table 4) of the terms assigned by the local libraries and not used by DDC appeared in one of the DDC vocabularies, DRIT (DDC Retrieval and Indexing Terminology) or CFC (Combined Frequency Count).

Table 4 SUBJECT INDEXING: COMPARISON OF DOD (ARMY, NAVY AND AIR FORCE) LIBRARIES AND DDC.

	<u>Navy</u>	<u>Army</u>	<u>Air Force</u>	<u>DOD</u>
Percentage of terms assigned by DOD libraries and <u>not</u> used by DDC...	49%	50%	34%	44.3%
Percentage of terms assigned by DOD libraries and <u>not</u> used by DDC but appearing in DDC vocabulary DRIT or CFC.....	39%	35%	20%	31.3%
Percentage of terms assigned by DOD libraries and <u>not</u> used by DDC and <u>not</u> appearing in any DDC vocabulary.....	10%	15%	14%	13%
Percentage of terms assigned by DOD libraries and appearing in DDC vocabulary DRIT or CFC (Includes terms used and not used by DDC for the indexing of items sampled).....	90%	85%	86%	87%

The breakdown of DRIT and CFC terms assigned by the DOD libraries but not used by the DDC analysts appears in Table 5.

Table 5 DISTRIBUTION OF DDC DRIT AND CFC TERMS USED BY DOD LIBRARIES BUT NOT USED BY DDC TO INDEX THE SAMPLED DOCUMENTS.

	<u>Navy</u>	<u>Army</u>	<u>Air Force</u>	<u>DOD</u>
DRIT	59%	48%	65%	57.3%
CFC	41%	52%	35%	42.7%

Results of Interview

The DDC thesaurus (DRIT/CFC) is used by 11 libraries (Table 6), and no library has any immediate plans to change the thesaurus it is presently using. A few librarians mentioned that their libraries recently changed the thesaurus they were using when they undertook the automation of their document retrieval system.

Table 6 THESAURI IN USE BY DOD LIBRARIES*

	<u>DDC (DRIT/CFC)</u>	<u>TEST</u>	<u>IN-HOUSE</u>	<u>OTHER</u>
Navy	2	7	5	
Army	6	3	3	2
Air Force	3	1	2	
DOD	11	11	10	2**

(*The reason the figures add up to more than 22 is that most libraries use more than one thesaurus or incorporate many of their in-house terms with one of the standard thesauri.)

(**AEC, NASA)

Except for the four libraries using DDC's cataloging, all of the libraries catalog documents produced in-house even if the documents are available at DDC. However, concerning documents produced out-of-house and received by the library, four libraries indicated they did not catalog unclassified microfiche documents from DDC. Two libraries do not catalog any microfiche documents if unassociated with a hard copy in the library. Thus some libraries are building microfiche collections which are accessible only by using DDC's "Technical Abstract Bulletin" and NTIS' "Government Research Abstracts & Index" or by going on-line to DDC, LOCKHEED or SDC. (Lockheed and SDC have the NTIS data base on-line.)

CONCLUSIONS

Following are conclusions drawn from the data obtained. Along with each conclusion are questions and suggestions which could lead to further research on the subject.

- a) DDC assigns more subject headings per document than almost all DOD libraries. Thus, DDC subject-indexes documents as exhaustively as do the DOD libraries - numerically speaking that is. But is more better? Does more bring a higher recall and precision rate? To the user, more is better if the key word(s) he or she has in mind can be found in or related to the indexing terminology needed to locate useful documents.
- b) Just a little over 50% of the terms assigned by DOD libraries are used by DDC analysts to catalog the same documents. Thus DDC does not subject-index documents as exhaustively as the DOD libraries, when comparing similar and identical terms. Yet 87% of the terms assigned by the DOD libraries are actually in the DDC vocabulary DRIT or CFC (Table 4). Why were not more of these terms assigned by DDC analysts to bring up the similar and identical count? Why does not DDC review the in-house cataloging of local documents to discern the major subjects relevant to that document as viewed by the catalogers of the organization producing the document? Why does not the local library input its cataloging into DDC via block 14 of the 1473 form? Two libraries interviewed do. Why not more?
- c) DDC and the DOD libraries are practically using the same vocabulary with which to index documents even though the thesaurus may not be the same. Perhaps DDC should prepare a better DRIT that would reflect

this and place those few classified documents in a separate vocabulary for many terms appear in CFC that should be in DRIT.

- d) Since DDC and the DOD Libraries have as one of their functions the indexing of documents to serve DOD user needs, why should they not share and cooperate in accomplishing this important task? DDC could index the report from a broad point of view while incorporating the local library's indexing to get the narrow or specific point of view. (Although from the sampled indexing examined, it was ironically often the other way around with DDC very specific and the local libraries indexing broad areas.) DDC should build upon local indexing to see that all aspects of a report are covered, be it broad or narrow. Accepting the local indexing outright would create more problems than it would solve for any semblance of standardization presently existing at DDC would be thrown out the window. DDC, like the Library of Congress, should be the standard bearer. DDC is presently investigating some aspects of shared cataloging via the on-line system and this might lead to the elimination or at least the diminution of the extensive duplicate cataloging and indexing going on throughout the country.

APPENDIXES

Appendix A SPECIFIC QUESTIONS ASKED DURING PHONE CONVERSATION.

1. Which thesaurus are you presently using to subject index the documents you catalog?
2. Are you planning to change the thesaurus you are now using in the near future?
3. Do you catalog your organization's documents if they are available from DDC?
4. Do you catalog documents from other organizations, if they are available from DDC?
5. Do you have an automated retrieval system for the documents you catalog? Yes___ No___

Appendix B ORGANIZATIONS WHOSE LIBRARIES WERE CHOSEN TO SUPPLY DATA AND INDEXING INFORMATION.

(*Organizations responded to phone interview but did not submit indexing information.)

Army *Army Air Mobility R&D Laboratory
 Ft. Eustis, Va.

 Army Aviation Systems Command
 St. Louis, Mo.

 *Army Mat. & Mech. Research Ctr.
 Watertown, Ma.

 Army Research & Development Command
 Natick, Ma.

 *Army Ballistic Research Lab.
 Aberdeen, Md.

 *Harry Diamond Lab.
 Adelphi, Md.

 Army Missile Command
 Redstone Arsenal, Ala.

(13)

Picatinny Arsenal
Dover, N.J.

Air Force Air Force Weapons Lab.
Kirtland AFB, N.M.

*Rome Air Development Ctr.
Griffin AFB, N.Y.

*Air Force Armament Test Lab.
Eglin AFB, FLA.

*Air Force Environmental & Tech. Applications Ctr.
Scott AFB, ILL.

Navy *Naval Air Development Ctr.
Warminster, PA.

Naval Coastal Systems Laboratory
Panama City, FLA.

Naval Electronics Laboratory Ctr.
San Diego, CA.

Naval Post Graduate School
Monterey, CA.

Naval Research Laboratory
Washington, DC.

Naval Surface Weapons Ctr. White Oak
Silver Spring, MD.

Naval Undersea Center
San Diego, CA.

Naval Underwater Systems Center
New London, CT.

Naval Underwater Systems Center
Newport, R.I.

Naval Weapons Center
China Lake, CA.

(14)

Author Ronald G. Heroux

Title Subject Indexing - DDC and the DOD Technical Libraries

Source Simmons College, Boston, MA.

Date Feb. 1977

Pages 13

Key Words libraries, technical libraries, information centers, *indexing, *subject indexing, technical reports, *cataloging, Defense Documentation Center, DRIT (DDC Retrieval and Indexing Terminology), information retrieval, documents.

The author makes the following

Abstract The subject indexing produced by the Defense Documentation Center and a randomly selected number of DOD technical libraries is examined to determine: (1) Whether DDC or the DOD technical libraries assign more subject indexing terms per document; (2) What percentage of subject indexing terms assigned by the DOD technical libraries are identical or similar to those used by DDC. and Conclusions: (1) DDC assigns more subject headings per document than almost all DOD libraries; (2) a little over 50% of the terms assigned by the DOD libraries are used by DDC analysts to catalog the same documents; (3) DDC and the DOD libraries are practically using the same vocabulary with which to index documents.

Distribution unclassified unlimited (available for public release.)